1/15/15 263 Lecture 1

Website: math.arizona.edu/~nhenscheid
Download & read Syllabus

Calendar:

Today: 1.1-1.3 Intro, Plots, 5# summary
Next Tue: 1.4 The Normal Dist
Next Thurs: 2.1-2.4, Relationships between Vars
Excel Assignment 1 DUE

Theme for Chapter 1: Understand your data set!

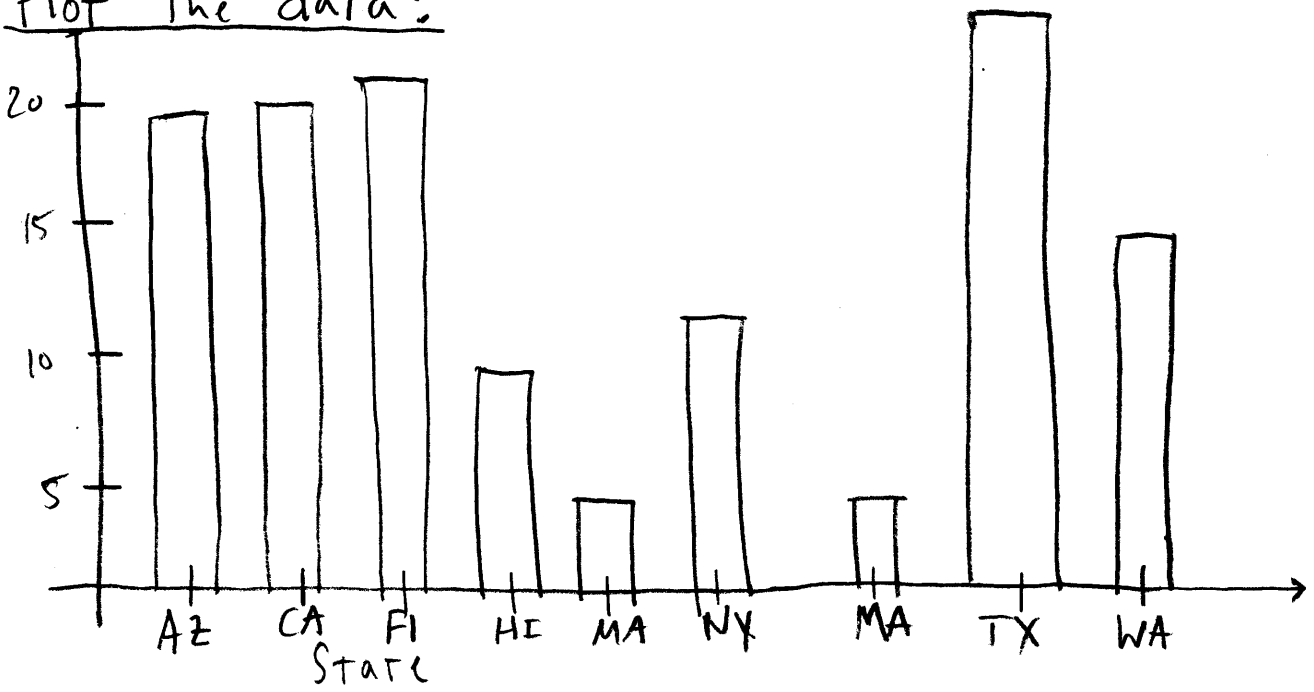
Example Health Insurance Coverage

State	% without insurance (2010-2012)
AZ	18.2
CA	19
FL	20.7
HI	7.8
MA	7.3
TX	24.3
WA	14
NY	12.9

* Cases: States
* Vars: State Name, Abbr. ("Label") & % w/out insurance.
Vars can be Categorical or Quantitative

Source: US Census Bureau

Plot the data:

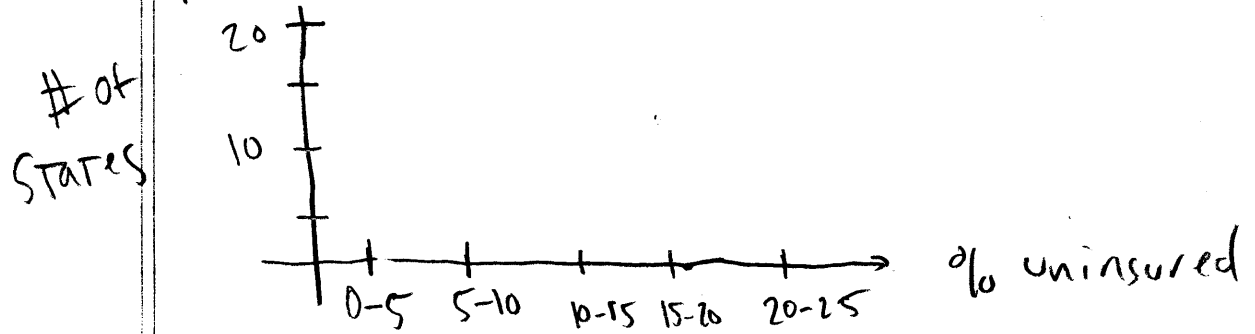


What does this represent?

- Helps identify <u>some</u> patterns & outliers
- Warning: Order is arbitrary! any trend or structure is meaningless.

*** We are more interested in the <u>distribution</u> of the data. If we choose another state @ random, what is <u>its</u> % uninsured?

Want a plot more like this:



A Distribution tells you what values a variable takes & how frequently it takes them.

Connection w/ Probability: If a certain value is taken less frequently, it is deemed "less likely" or "less probable"

Histogram: "X-axis" will have values in the range of a variable & "Y-axis" will have COUNTS

e.g. How many states have between 0 & 5 % uninsured?

IE. $\#\{0 \leq I < 5\}$ | (MA)

STATES

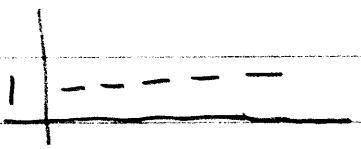
5 to 10%?

20-25%? | (TX)

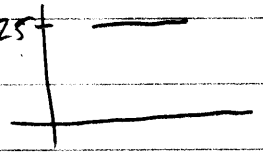
Bins	Counts
0-5	1
5-10	1
10-15	12
15-20	11
20-25	2

Bin widths are important!

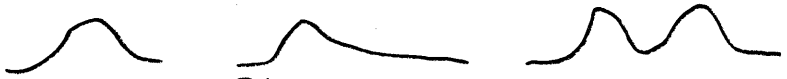
too narrow:



too wide:



Properties of Distributions:

- Pattern, Deviation
- Shape, Center, Spread
- Outlier
- modes 
- Symmetry vs Skew
  Is a small score as likely as a large score? Ex tests scores

State unisured: unimodal, ~~skewed right~~ ?

Quantifying Distribution Shape

- Want a numerical summary of distribution instead of all the data pts
  - We will cover the "5 # Summary" mean, median, Var & Std dev, Q1, Q3, IQR, min, Max
  - Box Plots 

("Sample") Mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ — each Val has same weight

* Note: Mean is Not Robust! →

Median: In a sorted list, Half above & half below "50%"

ODD:

1 3 7 [8] 12 15 16 $n = 7$

$\frac{n+1}{2}$

Even:

12 7 8 10 12

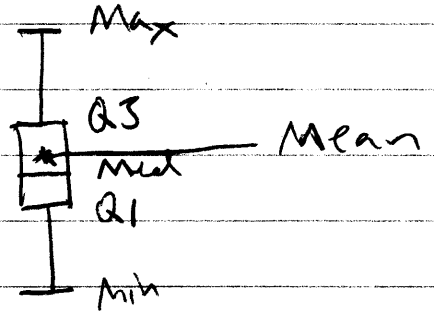
Average These 2

Q1: Median of Left Half
Q3: Median of Right Half
IQR: Q3 − Q1

* Note: Quartiles are Robust

5 # Summary: Min, Max, Median, Q_1, Q_2

Box Plot:



Outliers: 1.5 · IQR From Q1 or Q3
↝ Possible outlier.